3/22/06

Towards the characterization of upper tropospheric clouds using AIRS and MLS observations

Brian H. Kahn[1], Annmarie Eldering[1], Amy J. Braverman[1], Eric J. Fetzer[1], Jonathan H. Jiang[1], Evan Fishbein[1], and Dong L. Wu[1]

[1] Jet Propulsion Laboratory, Pasadena, CA, USA

Corresponding author and address:

Brian Hannon Kahn
Jet Propulsion Laboratory
4800 Oak Grove Drive
Mail Stop 169–237
Pasadena, CA, 91109
Fax: 818-393-4619
Tel: 818-393-0676
E-mail: brian.h.kahn@jpl.nasa.gov

Submitted to the *Journal of Geophysical Research-Atmospheres*

Last revised on March 22nd, 2006

Abstract

We estimate the accuracy of cloud top altitude (Z) retrievals from the Atmospheric Infrared Sounder (AIRS) and Advanced Microwave Sounding Unit (AMSU) observing suite (Z_A) onboard the Earth Observing System (EOS) Aqua platform. We compare Z_A with coincident measurements of Z derived from the micropulse lidar and millimeter-wave cloud radar at the Atmospheric Radiation Measurement (ARM) program sites of Nauru and Manus Islands (Z_{ARM}), and the inferred Z from vertically resolved Microwave Limb Sounder (MLS) ice water content (IWC) retrievals. The mean difference in $Z_A - Z_{ARM} \pm$ one standard deviation ranges from –2.2 to 1.6 km ± 1.0 to 4.2 km for all cases of AIRS effective cloud fraction (f_A) > 0.15 at Manus Island using the cloud radar only. The range of mean values results from using different approaches to determine Z_{ARM}, day/night differences, and the magnitude of f_A; the variation about the mean decreases for increasing values of f_A. Analysis of Z_{ARM} from the micropulse lidar at Nauru Island for cases restricted to $0.05 \le f_A \le 0.15$ indicates a statistically significant improvement in $Z_A - Z_{ARM}$ over the cloud radar-derived values at Manus Island. In these cases the $Z_A - Z_{ARM}$ difference is –1.1 to 2.1 km ± 3.0 to 4.5 km. These results imply that the operational Z_A is quantitatively useful for constraining cirrus altitude despite the nominal 45 km horizontal resolution. Mean differences of cloud top pressure (P_{CLD}) inferred from coincident AIRS and MLS IWC retrievals depend upon the method of defining AIRS P_{CLD} (as with the ARM comparisons) over the MLS spatial scale, the peak altitude and maximum value of MLS IWC, and f_A. AIRS and MLS yield similar vertical frequency distributions when comparisons are limited to $f_A > 0.1$ and IWC > 1.0 mg m^{-3}. Therefore, the agreement depends upon the opacity of the cloud, with decreased agreement for optically tenuous clouds. Further,

the mean difference and standard deviation of AIRS and MLS P_{CLD} are highly dependent on the MLS tangent altitude. For MLS tangent altitudes greater than 146 hPa, the strength of the limb technique, the disagreement becomes statistically significant. This implies that AIRS and MLS "agree" in a statistical sense at lower tangent altitudes, and "disagree" at higher tangent altitudes. These results provide important insights on upper tropospheric cloudiness as observed by nadir-viewing AIRS and limb-viewing MLS.

1. Introduction

Validating satellite cloud observations is made challenging by spatial and temporal cloud variability at scales similar to and smaller than the satellite measurement. Historically, satellite derived cloud quantities have been compared to ground-based measurements. Recently, a set of satellite instruments flying in formation, dubbed the "A-Train", provides a unique opportunity to cross-validate cloud measurements made from different angles and platforms with close coincidence. A high quality record of upper tropospheric cloudiness will lead towards an improved understanding of atmospheric hydrological processes, atmospheric transport in the tropopause region, and the role of cloudiness in Earth's present and future climate [*Houghton et al*. 2001].

Much effort over the years has been spent on the cross-comparison of cloud observations from *in situ*, airborne, spaceborne and surface-based instrumentation. This includes inter-comparisons of satellite-retrieved cloud properties to surface-based active measurements [e.g., *Smith and Platt*, 1978; *Vaubauce et al*. 2003; *Berendes et al*. 2004; *Hollars et al.* 2004; *Naud et al.* 2004; *Elourghani et al.* 2005; *Hawkinson et al.* 2005; *Mace et al.* 2005], retrievals from

multiple satellite platforms [e.g., *Stubenrauch et al*., 1999, 2005; *Wylie and Wang*, 1999; *Naud et al*. 2002; *Mahesh et al*. 2004], satellite to aircraft platforms [e.g., *Sherwood et al*., 2004], and multiple instruments collocated on single and multiple aircraft platforms [e.g., *Smith and Frey* 1990; *Frey et al.* 1999; *Holz et al.* 2006].

In situ measurements of cloud vertical location, particle size distribution, ice particle habit, extinction coefficient, water phase, and ice and liquid water content are critically important for independently validating remote sensing-derived cloud quantities [*Heymsfield and McFarquhar* 2002]. However, cross–comparing satellite-derived cloud quantities is equally important for assessing consistency over global atmospheric conditions, as well as for algorithmic and product improvements. The cross-comparison of cloud quantities is a more difficult problem than the comparison of other fields, such as temperature and water vapor [*Fetzer et al.* 2006; *Gettelman et al*. 2006], because of the greater variability of clouds on short temporal and spatial scales.

One of the more important physical quantities required for cloud-related research is cloud height [*Cooper et al.* 2003]. Surface-based active measurements, including the millimeter-wave cloud radar and micropulse lidar located at the Atmospheric Radiation Measurement (ARM) program sites [*Ackerman and Stokes* 2003], the Geoscience Laser Altimeter System (GLAS) located onboard the Ice, Cloud, and Land Elevation Satellite (ICESat) [*Spinhirne et al*. 2005], and the upcoming CloudSat [*Stephens et al.* 2002] and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) [*Winker et al.* 2002] missions, provide accurate and precise cloud base and height information across a wide-range of cloud optical thickness, but at horizontal scales of a few hundred meters and at a select number of locations, usually land-based. Satellite-based observations measure cloud properties at scales from hundreds of meters

to hundreds of kilometers and have daily global coverage, but are more uncertain. Aircraft-based observations lie in between while *in situ* measurements are the most accurate and precise, but have the smallest spatial and temporal resolution; these data will not be used in this analysis.

Among the satellite-based cloud height observations, those derived from passive infrared (IR) are the more uncertain, but have the greatest global coverage and longest time span. These satellite products must be compared with more precise and accurate active lidar and radar-derived measurements for non-opaque clouds. In addition, the Aqua satellite provides a unique opportunity to inter-compare simultaneous satellite-base cloud products from an IR imager, the Moderate Resolution Imaging Spectroradiometer (MODIS) [*King et al.* 1992], and an IR/microwave (MW) sounder suite, the Atmospheric Infrared Sounder (AIRS) [*Aumann et al.* 2003]/Advanced Microwave Sounding Unit (AMSU) with ground-based upward-looking radar and lidar measurements at the Tropical West Pacific ARM sites. This comparison is especially important because the Visible Infrared Imager/Radiometer Suite (VIIRS) on future operational platforms, i.e. the National Polar-orbiting Operational Environmental Satellite System (NPOESS) and NPOESS Preparatory Program (NPP), will not have the MODIS 15 μm channels, and cloud height products may have to be derived in combination with the onboard Cross-track Infrared Sounder (CrIS)/Advanced Technology Microwave Sounder (ATMS) Sounder Suite (CrIMMS) [*Cunningham and Haas* 2004].

Cloud height is derived from IR CO_2 temperature sounding channels, used in a technique called "CO_2 slicing" [*Smith and Platt*, 1978 and references therein; *Menzel et al*. 1983] where calculated radiances from an operational weather forecast model are differenced with observed radiances, and the height is derived from ratios of the differences. The largest error sources from CO_2 slicing include those from the calculated radiances that have error contributions from

spectroscopic parameters, and others that are needed to calculate radiances, e.g., surface emissivity. These errors are especially large near the surface but could affect cloud products several kilometers above the surface because of the relatively broad weighting functions of the MODIS temperature-sounding channels. In general, the CO_2 slicing method is most accurate for high and opaque clouds, with increasing degradation in accuracy as cloud height and opacity decrease [*Wielicki and Coakley* 1981]. Other studies have shown that an improvement in CO_2 slicing-derived Z is observed when accounting for surface emissivity effects over land [*Zhang and Menzel* 2001]; however, multilevel cloud cover was shown [*Baum and Wielicki* 1994] to introduce errors in Z. More recent work by *Holz et al*. [2006] suggests that sorting CO_2 slicing channels by optical depth to determine optimal channel pairs using high-spectral-resolution measurements improves Z retrievals for tenuous cirrus and reduces errors introduced by surface and atmospheric uncertainties.

AIRS/AMSU uses a different approach called cloud clearing [*Chahine* 1974; *Susskind et al.* 2003]. The algorithms implemented in the version 4 production software uses MW sounding-derived temperature and water vapor profiles to predict clear sky IR radiances for a select set of channels. A linear combination of radiances from a 3×3 array of adjacent AIRS footprints is used to infer the clear-sky radiance over the entire AIRS spectral range. The cloud-cleared AIRS radiances are used to retrieve T(z), q(z), O_3(z), additional trace gases, and other atmospheric and surface properties. The cloud top pressure (P_A) and effective cloud fraction (*f*) are derived by comparison of observed AIRS radiances to calculated ones. The major sources of error in cloud-clearing arise from errors in the temperature and water vapor profiles provided by AMSU, especially near the surface, and uncertainty in calculated radiances from surface sensing channels, especially over land. Singularities in the cloud-clearing algorithms occur when clouds

are uniform across the 3×3 array of AIRS footprints. The first source of error is especially problematic because of the failure of the MW-based Humidity Sounder from Brazil (HSB) and interference in two of the AMSU temperature sounding channels (channel 7 is unusable).

None of the results presented in this study use the HSB radiances, but *Fetzer et al.* [2006] have shown that biases of retrieved total precipitable water vapor are less than 5% before and after the loss of HSB in the tropics and subtropics (the area of focus in this study) outside of the dominant stratus regions. However, precision was lost with HSB, and its impact on retrieved quantities (including clouds) is an ongoing subject of research. The effective cloud fractions have an additional error arising from assumptions about the radiative properties of the clouds, that they are opaque and spectrally black (IR emissivity = 1). In the full IR/MW retrieval, up to two levels of cloud top pressure (P_A), cloud top temperature (T_A), and nine effective cloud fractions (f_A) per cloud layer are retrieved. If the retrieval defaults to a MW-only retrieval, only one f_A is reported for each cloud layer. In this work P_A is converted to Z_A at the ARM tropical Western Pacific (TWP) sites using a typical TWP atmospheric profile with a scale height of 8 km.

Since the launch of EOS Aura on July 15, 2004, the Microwave Limb Sounder (MLS) has joined the A-Train and made global, vertically-resolved measurements for a wide array of gaseous species, along with temperature, geopotential height, and cloud ice water content (IWC). These observations are relevant for understanding stratospheric ozone chemistry, climate, and air pollution processes [*Waters et al*. 2006]. The MLS observes limb thermal emission radiation at millimeter- and sub-millimeter-wavelengths at a spatial resolution of approximately 200 km × 7 km × 3 km (along-track × cross-track × vertical); a limb scan is made every 165 km along track for a global total of ~3500 day^{-1}. The MLS is designed to retrieval geophysical parameters (temperature and a number of chemical compositions) from measured spectral line radiance

features, based on optimal estimation [*Livesey et al*. 2006] with Tikhonov regularization applied. The retrieved parameters, or state vectors, consist of vertical profiles on fixed pressure surfaces having a semi-global (82°N–82°S) coverage. MLS cloud measurements are divided into two steps. First, cloud induced radiances (ΔT_{cir}) are obtained by calculating the difference between the measured radiance and modeled clear-sky radiance using the retrieved atmospheric state. Second, high altitude (215–68 hPa) cloud ice water contents (IWC) are retrieved from ΔT_{cir} using modeled ΔT_{cir}–IWC relations [*Wu el at.* 2006]. The IWC data used in this research are from the MLS IWC version 1.51-CLD02 dataset described by *Jiang et al* [2006]. The vertical profile of IWCs are represented by equally-spaced increments in log-pressure, i.e. at 215, 178, 147, 121, 100, 83 and 68 hPa pressure levels.

The primary objective of this study is to assess how well AIRS and MLS determine cloud altitude in the upper troposphere, emphasizing the tropical Western Pacific warm pool region. In Section 2 we point out the complexities of comparing surface-based ARM point measurements to those made from satellite platforms like AIRS. The averaging methods of the ARM observations for replicating the horizontal scale of the satellite measurement are discussed. In Section 3 the AIRS and ARM-derived Z are compared, exploiting the different sensitivities of the micropulse lidar and mm-wave cloud radar to thin and thick cloud. The statistical significance of the results is addressed, and comparisons are made to previously published work. In Section 4 we compare AIRS and MLS-derived P_{CLD}, and show the agreement is conditional upon the opacity of the cloud, as well as the MLS tangent altitude. In Section 5 we summarize the results.

2. ARM and AIRS intercomparisons

In this work we examine the uncertainties in the retrieved Z_A, emphasizing the upper level Z_A because of its relevance to upper tropospheric cloud cover, using the active sensor measurements at the ARM TWP program sites at Manus and Nauru Islands, located at 2°S 147.5°E, and 0.5°S 167°E, respectively [*Ackerman and Stokes* 2003; *Mather* 2005]. AIRS is compared to the Active Remotely-Sensed Cloud Locations (ARSCL) value-added product (VAP) [*Clothiaux et al.* 2000]. The ARSCL VAP combines information from micropulse lidars, microwave radiometers, millimeter-wave cloud radars, and laser ceilometers into a time series of cloud tops and bases, allowing for a comprehensive database throughout the range of cloud height in the troposphere, optical depth (τ), and hydrometeor characteristics. For the time period of coincident measurements made by AIRS and the ARM sites (April–September 2003), the micropulse lidar was not operational at Manus Island; likewise, the mm-wave cloud radar was not operational at Nauru Island. As will be shown below, comparing Z_A to the Z_{ARM} derived from the two ARM instruments indicates the strength of Z_A not only for optically thick cloud, but optically thin cloud ($f_A < 0.15$).

In this study we compare an essentially instantaneous, downward–looking, passive measurement of clouds from space at the horizontal scale of 45 km and greater with an active upward–looking, surface-based, point measurement [*Kahn et al*., 2005]. Much of the discrepancy between independent measurements of Z is attributable to the fundamental heterogeneity of cloud properties over the scenes compared [see *Stubenrauch et al*. 1999]. If cloud spatial heterogeneity effects are to be eliminated in a comparison of two independent observations, homogeneous and static cloud properties over the spatial and temporal scales spanned by the two observations are necessary. However, observed cloud fields rarely provide

such opportunities [*Horvath and Davies* 2004; *Chylek and Borel* 2004]; thus comparisons must be made with less idealized cases.

A useful comparison approach is to consider the surface point observations over a period of time assuming a mean wind speed, giving an advective spatial scale of the satellite footprint. A discussion on the pitfalls in such an analysis can be found in *Kahn et al.* [2005]; these are related to cloud heterogeneity coupling to 3-D wind speed and direction gradients, and the evolving nature of the physical properties of clouds. Clouds are highly heterogeneous over the ~45 km scale of the AMSU footprint considered here [see *Cahalan et al.* 1989]. The satellite pixel is rarely centered on the surface point observation, further complicating the comparison. Additional instrument and algorithm-related sources of discrepancy are mentioned briefly in a companion paper by *Kahn et al*. [2006], and are discussed in detail by *Cracknell* [1998].

Given the complexity of cloud fields, we examine several ways to estimate the spatial mean from time-varying observations at a point. Taking into account the variability of wind speed, three different time-averaging procedures were applied to the ARSCL data: (1) Z_{ARM} is averaged over the 6 min period of the coincident AIRS granule ± 24 min for a total of 54 min, (2) ± 60 min for a total of 126 min, and (3) ± 90 min for a total of 186 min. The method used here does not take into account the wind speeds for individual cases. For the nominal AMSU footprint of 45 km, the three time averages correspond to wind speeds of 12.3, 5.3, and 3.6 ms^{-1}, respectively. For examples of using wind profiler measurements in individual comparisons of Z_{ARM} and MODIS-derived Z see *Mace et al.* [2005].

The Z_{ARM} within each time window is derived three different ways: (1) by an average of the highest ARSCL cloud top values for each observation time, ignoring clear sky (Z_{ARM}^{AVG}), (2) by developing histograms of all ARSCL cloud top values in 0.5 km bins within the time window,

then choosing the highest peak of "significance" in the distribution (Z_{ARM}^{HIST}), and (3) by choosing the highest value of cloud top height in the time window (Z_{ARM}^{MAX}). For Z_{ARM}^{HIST}, we define the height by: (*i*) assigning the ARM Z observation to height bins as specified in (2), (*ii*) locating all peaks in the frequency of occurrence, and (*iii*) then identify the peak with the highest altitude. If the number of cases in the highest peak is greater than approximately 10% of the value of the maximum peak (in the case of multiple peaks), it is used as Z_{ARM}^{HIST}. Otherwise the next highest peak with the greatest frequency of occurrence is used.

Since Z_{AIRS} is representative of a "cloud top", using the highest peak is physically justified. However, Z_{ARM}^{HIST} may be affected significantly by broken cloud cover within the AMSU FOV; the vertical cross-section of ARSCL observations is not necessarily a representative sub-sample for the entire AMSU FOV. Additionally, satellite CO_2 slicing-derived Z for deep and tenuous cirrus layers often place Z well below the Z observed by lidar [*Holz et al.* 2006]. Sampling errors using ARM measurements are essentially random, while errors due to limitations in the CO_2 slicing method Z are systematic. Therefore, the errors that impact comparisons of Z_{AIRS} and Z_{ARM} are a combination of random and systematic sources.

Next, we compare AIRS and ARM-derived Z for Manus and Nauru Islands. For the time period considered (April–September 2003) the lidar was not in operation at Manus; likewise the cloud lidar was not in operation at Nauru. As a result, optically thick clouds are better represented at Manus, and tenuous cirrus is better characterized at Nauru. Then we address the statistical significance of the results and compare them to other cloud height datasets.

3. Results

a. Manus Island

Results for the time average of ± 24 min and for the three Z_{ARM} methods for nighttime at Manus Island are presented in Figure 1, and are summarized in four main points. First, note Z_{ARM}^{AVG} disagrees for the high and optically thick clouds (large circles). This effect has been discussed elsewhere and is an apparent consequence of the attenuation of the millimeter-wave cloud radar beam due to precipitation [*Hollars et al.* 2004]. In the Z_{ARM}^{HIST} and Z_{ARM}^{MAX} cases this effect is reduced as expected since they are less affected by occasional precipitation-induced outliers. Second, in the Z_{ARM}^{AVG} case, many of the high and optically thin clouds are biased low compared to Z_A. In the Z_{ARM}^{AVG} approach all clouds detected in the ARSCL product are used, including low-level trade wind cumulus. Thus, a low bias is expected because there are instances in a time window when trade wind cumulus is the highest cloud. The Z_{ARM}^{HIST} and Z_{ARM}^{MAX} approaches show many of these cases to be in better agreement. Since the ARM measurement samples a point in horizontal space and a vertical plane over time (assuming a constant wind speed), broken cloud scenes may not be observed by ARM when detected in the AMSU FOV, further explaining why some cirrus cases are not observed in the ARSCL product.

Third, note the two clusters of cloud: one near 6–8 km, and the second from 9–15 km. They are consistent with the altitudes of peak frequency of tropical clouds [*Comstock and Jakob* 2004; *Hollars et al.* 2004]. Fourth, Z_{ARM}^{MAX} is biased high for clouds in the 9–12 km range. This effect is seen in an independent comparison of GOES Z with active lidar and cloud radar measurements [*Hawkinson et al*. 2005]. Many of these highest clouds have *f* substantially < 1 and are in better agreement in the Z_{ARM}^{HIST} case. This supports the notion that AIRS may not

always see the top of the highest cloud in a scene that is broken, multi-layered, or semi-transparent, but may place Z_{AIRS} somewhat lower [*Sherwood et al.* 2004; *Holz et al*. 2006].

The vertical and horizontal bars in Figure 1 for Z_A and Z_{ARM} represent two types of variabilities. For AIRS (the vertical bars), they are the 1–σ variability in Z_A from a plane-parallel radiative transfer calculation, taken from the AIRS L2 Standard product; these values tend to be inversely proportional of f_A [*Wielicki and Coakley* 1981]. For ARSCL (the horizontal bars) they represent the 1–σ variability in Z_{ARM} within a given time window. Essentially, the length of the bars on Z_{ARM} is a measure of the heterogeneity of Z_{ARM}. As a result, we do not expect the uncertainties to intersect the 1–1 line for every case, as is seen in Figure 1. On the other hand, there is a tendency for the uncertainties to be smaller when the agreement in Z is closer.

The distribution of the differences between Z_A and Z_{ARM} is shown in Figure 2. The agreement for the three time windows used is nearly constant, suggesting little sensitivity to the choice of the time window used to calculate Z_{ARM} at Manus. In the Z_{ARM}^{AVG} approach, the peak frequency in the agreement is located near AIRS–ARM ≈ 0 km, with a tail in the distribution for AIRS–ARM > 0; this effect is primarily caused by "missed detections" by the ARM measurements, as well as occasional frequent occurrences of cumulus clouds that bias Z_{ARM}^{AVG} low, as shown by visual inspection. The Z_{ARM}^{HIST} approach has a peak frequency near AIRS–ARM ≈ –1 km. The Z_{ARM}^{MAX} approach has a similar peak but a slightly more narrow distribution than Z_{ARM}^{AVG} and Z_{ARM}^{HIST}.

Table 1 summarizes the bias, standard deviation (1–σ), and statistical significance of the difference between coincident AIRS and ARM-derived cloud height. The bias and 1–σ of the

difference in Z_A and Z_{ARM} are a strong function of f_A. For the nighttime Manus Island cases, the bias ranges from several kilometers for $f_A < 0.15$ to less than 1 km for the optically thicker clouds. In the case of $f_A < 0.15$, much of the bias is likely explained by missed cirrus detections at the Manus site using the cloud radar. Within each category of f_A, the bias can vary by about 1–2 km when using different techniques to define Z_{ARM}. As in the case of the bias, the largest 1–σ variability is found for the thinnest clouds, and reduces to 1–2 km for the thicker clouds. This is likely related in part to the improved sensitivity of CO_2-slicing in higher and optically thicker clouds, and a more spatially uniform cloud height for deep convective cloud systems.

For the nighttime Manus Island cases shown in Table 1, all Z_{ARM} averaging methods for $f_A > 0.5$ agree at the 0% significance level using a *p*-test for statistical significance (see the Appendix for a detailed discussion of the *p*-test). For an arbitrary value of significance, say 5%, this indicates that less than 5% of thousands of "randomized" pairs explain the observed agreement. For $0.15 < f_A < 0.5$, all cases agree at the 5% significance level or less, although the Z_{ARM}^{HIST} cases agree at the 1% significance level. This may indicate the histogram approach is most appropriate for tenuous cirrus, since AIRS is sensitive to thin cirrus altitude. For $f_A < 0.15$, all of the cases fail the *p*-test, except for the two longest time windows using Z_{ARM}^{AVG}, which pass at the 5% significance level. The results of the *p*-test are consistent with the expected accuracy of Z_A.

Figure 3 shows the agreement for the daytime coincident measurements. Fewer cases of cloud cover near 6–8 km are noted, and there is a higher number of opaque clouds than at nighttime. Many of the issues described in Figure 1 apply to Figure 3: a few deep convective clouds agree better for Z_{ARM}^{HIST} when compared to Z_{ARM}^{AVG}, some of the thin cirrus cases come into better agreement for Z_{ARM}^{HIST} and Z_{ARM}^{MAX} when compared to Z_{ARM}^{AVG}, and the level of disagreement is

seen to increase slightly from 9–12 km for Z_{ARM}^{MAX}. Note the two outliers in the lower right of each panel in Figure 3; the AIRS/AMSU microwave-only retrievals are contaminated by precipitation in these cases, confirmed by inspection of the ARSCL data. This demonstrates that caution must be used when using Z_A in precipitating clouds.

In Figure 4 the distribution of agreement is qualitatively similar to Figure 2, although there is a reduction in the difference between the different methods of Z_{ARM}^{HIST} compared to Figure 2. Additionally, the width of the distribution in Figure 4 is slightly larger than that of Figure 2, implying slightly worse agreement between AIRS and ARSCL during daytime; this behavior is noted in the discussion of Table 1 as well.

As a whole, the daytime and nighttime cases at Manus compare reasonably well to each other in Figure 4 and Table 1 with regard to bias, 1–σ, and statistical significance tests, with a few notable exceptions. For $f_A > 0.85$, the agreement is not significant at the 0% level and, for the two longer Z_{ARM}^{HIST} time windows, they do not pass the 5% level. If the two outlying AIRS retrievals in Figure 3 are removed, all cases for $f_A > 0.85$ pass at the 0% significance level (shown in Table 1). The two outliers are contaminated by precipitation, which is confirmed upon inspection of the Manus ARSCL product. The AIRS L2 quality flags are discussed in *Susskind et al.* [2006]. However, they are not used in this work to remove precipitating clouds since we are validating the entire range of cloud opacity. At nighttime two of the Z_{ARM}^{AVG} cases for $0.05 < f_A < 0.15$ pass at the 5% significance level; for the daytime cases, this is no longer the case. The bias and variability increase substantially for this range of f_A in the daytime compared to nighttime. Diurnal effects can be attributed to the small sample size, to differences in diurnal cloud amount, or systematic diurnal biases in the retrieval algorithms. An extended period of coincident observations will help address this question, and is the subject of future research.

The results for Manus Island summarized in Table 1 indicate that the agreement there in Z for the thin cirrus clouds ($f_A < 0.15$) is no better or, perhaps worse, than what would be expected of randomness. The cloud radar does not detect much of the high and thin cirrus that is ubiquitous throughout the tropical upper troposphere [*Comstock et al.* 2002]. As a result we expect that the $f_A < 0.15$ cases at Manus fail the *p*-test. During April–September 2003 the micropulse lidar was in operation at Nauru Island (with the cloud radar inactive), and is the basis of Z_{ARM}. We use the coincident AIRS and ARM measurements during this time period to explore the accuracy of Z_A for thin cirrus clouds with $f_A \leq 0.15$.

b. Nauru Island

In Figure 5 we show Z_A and Z_{ARM} for Nauru at nighttime. One of the most obvious differences between Figures 1, 3, and 5 is the size of the estimated horizontal and vertical standard deviations. In the case of Z_{ARM}, the lidar appears to capture more layers of cloud than the radar alone in the $f_A \leq 0.15$ cases. In addition, the frequent occurrence of near-surface cumulus clouds helps to increase the size of the horizontal bars emphasizing ARM variability. For Z_A the variability is similar to the cases of $f_A < 0.15$ in Figures 1 and 3. As a whole, the variability decreases as f_A increases. In Figure 5 the results for Z_{ARM}^{HIST} is shown using the three time window averages. The agreement depends on the length of the time window, unlike what is observed at Manus. A higher number of the thinnest cirrus cases come into agreement (within the 1-σ bars) as the time window lengthens, and many fewer "missed" detections by the ARM site are noted for the longest time window.

Another interesting feature shown by Figure 6 is that the peak in Z_A is roughly two km lower than the Z_{ARM} peak. This is apparently due to the prevalence of very thin and patchy cirrus, some with multiple layers, at Nauru. The AIRS retrieval therefore places Z_A lower than what is seen at Nauru. This altitude can be substantially below the cloud top in tenuous cirrus clouds. Note the change in the behavior of the distribution as the time window increases (upper panel of Figure 6). As the length of the time window increases, the distribution narrows, and the peak near $Z_A - Z_{ARM} = 7$ decreases in magnitude. Most of the cases that cause the change in the distribution are for $f_A < 0.05$. In the discussion of Table 1 above, we note that the statistical significance of the agreement for these cases fails the *p*-test. However, as the time window increases in length, many cases have improved agreement with a lower bias; this behavior is consistent with the patchy nature of cirrus in these cases at low f_A. A larger time window is more likely to capture a portion of thin cirrus. It should be pointed out that $f_A = 0.05$ should not necessarily be used as an absolute "cut-off" for validated retrievals. It is possible that a threshold for a statistically significant agreement at the 0–5% level could be found at a lower value of f_A.

With respect to the statistical significance of $Z_A - Z_{ARM}$ at Nauru Island, the agreement is significant at the 0% level for the cases with $0.05 < f_A < 0.15$, shown in Table 1. This is a vast improvement over the statistical significance at Manus Island for the same range of f_A. The bias and 1-σ variability is approximately –1.1 to 2.1 km ± 3.0 to 4.5 km; this compares to –1.5 to 6.3 ± 3.0 to 8.5 km at Manus. Recall from Table 1 that the range of values results from using different approaches to determine Z_{ARM} and the magnitude of f_A. We attribute the better agreement at Nauru to the differences in the sensitivity of the lidar to thin cirrus clouds, consistent with previously published results [*Comstock et al.* 2002]. The Nauru comparison highlights the sensitivity of the AIRS upper level cloud top pressure product to tenuous cirrus

clouds. In future work, we will identify a much larger set of cases as more AIRS and ARM collocated observations are made over time. A statistically meaningful sampling rate requires a larger set of cases in order to reduce the width of the f_A bins and keep the number of cases in each bin large enough so that a representative sub-sample of observed cloud variability is obtained.

We summarize the results at Manus and Nauru Islands in Figure 7. The frequency distribution of $Z_A - Z_{ARM}$ for the radar cases at Manus compare favorably to the results of *Hollars et al*. [2004], who relate GMS–5 geostationary satellite-derived Z to Z_{ARM} at Manus. The shapes of the distributions are very similar, including the tails, although the peak frequency is quite sensitive to the method of defining Z_{ARM}. At Nauru the bias in peak frequency in this study is higher than *Frey et al*. [1999] and *Hawkinson et al*. [2005]. In *Frey et al*. [1999] the MODIS Airborne Simulator (MAS) is compared to coincident observations from a lidar on the same aircraft platform. Since both are downward-looking instruments, different sensitivities are expected to those from upward-looking active instruments, as are the ARM sites. Additionally, the pixel size of comparison in *Frey et al.* [1999] is aggregated to roughly 2 km, which is much smaller than the nominal footprint size of 45 km considered here. In *Hawkinson et al*. [2005], Z is derived from Geostationary Operational Environmental Satellite (GOES) radiances and is compared to a combination of radar– and lidar–derived Z. As in the case of *Frey et al*. [1999] and *Hollars et al*. [2004], the field of view of comparison is much smaller than AIRS. Lastly, it should be reiterated that the cases shown here are for $f_A < 0.15$; in the other comparisons the full range of cloud opacity is compared. The most tenuous cirrus cases are expected to have much poorer agreement than more opaque cloud cover (as shown in Table 1); *Frey et al.* [1999] and

Hawkinson et al. [2005] can be expected to have better agreement than the results presented here for Nauru.

4. Comparing AIRS and MLS cloud properties

a. Different perspectives of coincident cloud fields

The eight minute difference in measurements of clouds by the AIRS and MLS instruments provide a unique opportunity to cross-examine AIRS cloud height and fraction fields with a large number of near-simultaneous coincident MLS IWC measurements. This offers the potential of characterizing the 3-D spatial state of high-level cloudiness [*Liou et al*. 2002]. Nadir IR sounding instruments, such as AIRS, have limitations in resolving vertical cloud structure but have good horizontal resolution. Conversely, limb instruments like the MLS have limitations in their horizontal field of view, but are ideal for making measurements of clouds at fine vertical resolution. The two very different cloud measurements can be compared if the limitations and strengths of each measurement are understood in terms of the sensitivity to particular cloud systems.

AIRS and MLS have numerous differences in their instrumentation, measurement wavelengths, retrieval techniques, and viewing geometry. The MLS senses through clouds horizontally in thin vertical layers, while AIRS observes radiance from a vertical or near-vertical atmospheric column. Because the effective size, size distribution, and habit of ice particles significantly attenuate radiance as a function of wavelength, the AIRS and MLS instruments effectively observe different parts of the same cloud. Because MLS views in the forward along-

track direction, with the tangent altitude trailing the AIRS nadir observations by approximately 8 min, the instrument, wavelength, retrieval technique, and viewing geometry-related differences likely dominate over those due to cloud evolution. However, cloud evolution should not be dismissed, especially in rapidly evolving convective systems. Vertical velocities greater than 10 ms^{-1} are not uncommon in convective systems; this translates to a distance > 4 km in an 8 min period.

MLS is sensitive in detecting ice particles from a few tens to a few hundreds of microns in diameter in optically dense clouds [*Wu et al.* 2006]. It is able to sense into and through many thick cumulonimbus clouds such as those associated with deep-convection and anvil outflow but can miss thin cirrus composed of smaller ice particles. The AIRS instrument is most sensitive to tenuous clouds but saturates around an IR optical depth above 5 [*Huang et al.* 2004]. A trade-off exists between the reduced sensitivity of MLS to small particle ice clouds compared to AIRS, and the longer limb-viewing path length of MLS (about an order of magnitude longer than the nadir view) [*Kahn et al*. 2002]. Also, the MLS reports IWC at discrete pressure levels, and AIRS reports cloud top pressure (P_A) at a continuous range of altitudes.

We compare the MLS IWC to P_A for each of the six pressure levels between 82 and 215 hPa. The best MLS sensitivity to cloud top height is at 100–147 hPa. MLS levels are discrete and P_A is continuous, necessitating a comparison on an MLS level-by-level basis. The comparisons are subcategorized as a function of the AIRS upper level cloud fraction (f_A) and MLS IWC. We define the MLS cloud top pressure (P_M) as the highest altitude (lowest pressure) with IWC > 0 mg/m^3. Some values of cloud-induced radiance are below a nominal clear-sky uncertainty; the sensitivity of the agreement to removing such cases will be discussed.

The AIRS P_A horizontal FOV is roughly circular at ~45 km, and the MLS FOV is 165 km × 7 km (along-track × cross-track), complicating comparisons in heterogeneous cloud cover. Thus we represent P_A in two ways: the three nearest P_A retrievals to the MLS 100 hPa tangent point along the line of sight are averaged (P_{AVG}), and the lowest value of P_A (highest altitude) among the three nearest P_A retrievals is set to P_{HI}. In the case of P_{AVG} the total horizontal AIRS FOV along the MLS line of sight is roughly 135 × 45 km (along-track × cross-track), which is considerably wider than MLS cross-track FOV. In the case of P_{HI} previous studies of clouds in occultation measurements [*Kahn et al*. 2002, and references therein] suggest the cloud may fill a very small portion of the FOV along the line of sight; this motivates the use of P_{HI} for determining P_A.

b. Results

The frequency distribution of P_A and P_M for collocated and coincident cloud observations is shown in Figure 8. In Figure 8a, the P_{AVG} indicates a peak cloud occurrence near 200 hPa, with a rapidly decreasing frequency at higher and lower pressures. In contrast, the P_{HI} shows a peak closer to 170 hPa, and a secondary peak near 90–100 hPa; the secondary peak is not observed in the P_{AVG} approach, as it appears to smooth over the highest altitude cloud retrievals. The secondary peak in P_{HI} is dominated by clouds where $f_A < 0.1$. It resides around the altitude where an extensive layer of geometrically and optically thin cirrus is known to exist [*Peter et al.* 2004; *Dessler et al*. 2006]; however some of these values could be spurious. The CALIPSO lidar [*Winker et al*. 2002] is expected to be useful in quantifying the sensitivity of AIRS

retrievals to thin cirrus clouds near the tropopause because of the opportunity to collocate measurements within minutes of each other.

In Figure 8b the cloud occurrence frequency distribution for the MLS pressure levels is shown. When all cases are included the shape of the frequency distribution is substantially different than that for AIRS in Figure 8a, with a peak higher in altitude. If the MLS measurements in which the peak values of IWC < 1 mg m^{-3} within a given occultation are not considered, the shape of the curve remains similar, although the highest clouds are mostly removed. The frequency distribution changes substantially for the cases when the highest MLS level with IWC < 1.0 mg m^{-3} are excluded; the peak frequency occurs at a similar altitude as AIRS. A slightly larger number of cases occur above the peak for MLS, which is likely due to MLS finite (~3.5 km) vertical FOV. With this antenna pattern, MLS tends to report clouds at a slightly higher altitude where clouds only partially enter the FOV. Additionally, a larger number of MLS cases occur below the peak for MLS; this is because MLS reports cloud at all tangent altitudes, not the top of P_{CLD}, as in AIRS.

Figure 9 summarizes $P_A - P_M$ versus MLS pressure level for the cases presented in Figure 8. Considering the P_{AVG} and P_{HI} methods for "All Cases", the best agreement is for the 100–177.8 hPa levels, using the width of the distribution as a measure of agreement. The bias, 1–σ, and statistical significance for all MLS tangent altitudes using the P_{AVG} and P_{HI} methods are summarized in Table 2. P_{AVG} is biased low in pressure (high in altitude) relative to P_M at the lower tangent altitudes, transitioning to a high bias in pressure at the upper tangent altitudes. The 1–σ of the $P_{AVG} - P_M$ distribution reaches a minimum at P_M = 121.2 hPa, increasing at lower and higher tangent altitudes. The results for P_{HI} are similar except that the P_{HI} bias is consistently lower in pressure by 30–80 hPa than P_{AVG} (or 1–2 km), which is consistent with MLS half-power

beamwidth of 3.5 km Additionally, the 1–σ variability is consistently lower for P_{HI} compared to P_{AVG}. This suggests that a simple average of multiple AIRS P_{CLD} retrievals along the MLS line of sight is less representative of the actual cloud altitude than is the highest AIRS P_{CLD} when comparing to a limb sounding-derived cloud measurement [*Kahn et al*. 2002].

AIRS and MLS clouds show an improved agreement as IWC increases. Table 2 shows the bias, 1–σ and statistical significance for subsets of coincident retrievals when the highest in altitude IWC value in a given MLS occultation is greater than a specified threshold (the upper row of Table 2). The MLS IWC retrievals are inherently noisy at low values because of uncertainties in the clear-sky signature; this uncertainty is quantified at the 3–σ level as a function of MLS tangent altitude [*Wu and Jiang* 2004]. The AIRS–MLS P_{CLD} difference tends to decrease as higher IWC cases are considered, and the 1–σ values tend to decrease. The implication is that the AIRS and MLS P_{CLD} agrees better in the high IWC cases, when diffuse (low IWC) cloud top cases are not considered.

The statistical significance of the AIRS–MLS comparisons are highlighted in Table 2. This comparison is not analogous to that of AIRS to ARM measurements because MLS retrieves IWC at discrete levels. A randomization test like the one described in the Appendix for continuous quantities is not appropriate here. Here we test the hypothesis that P_M at a given level could be a plausible sample from the distribution provided by P_A at that level. To test this we group P_A according to classes partitioned by P_M. Then, we conduct a separate hypothesis test for each class. If P_M is greater than 97.5% or less than 2.5% of the P_A values, we can reject the null hypothesis, and infer that the two distributions of P_A and P_M are statistically different at the 5% level.

Table 2 shows that we cannot reject the hypothesis that P_M comes from a distribution of P_A for most MLS tangent altitudes. This implies that the AIRS and MLS P_{CLD} are statistically "similar" and could come from the same distribution of clouds. In other words, AIRS and MLS "agree" at these levels. However, this is not true at the highest altitudes, and especially for P_{AVG}. In these instances we can reject the hypothesis that AIRS and MLS P_{CLD} come from the same statistical distribution of clouds. In other words, AIRS and MLS "disagree" at these levels. The sample size varies with the IWC range considered, but the statistical significance is generally invariant with IWC (not shown). This result implies that AIRS and MLS are observing different aspects of the same cloud system at the highest MLS tangent altitudes, which is related in part to the geometry, wavelength, and retrieval method differences between the two instruments.

In summary, the agreement of AIRS and MLS depends on the method of determining AIRS P_{CLD}, the peak altitude of MLS IWC, the maximum IWC amount, and f_A. Some of the main findings are: (1) the vertical frequency distribution of P_M agrees well in cases where tenuous cirrus cases are not considered, namely, when the highest MLS tangent level has IWC > 1 mg m^{-3}, (2) the bias and 1–σ variability in AIRS and MLS P_{CLD} decreases for increasing MLS IWC, (3) P_{AVG} is biased high in altitude to P_M at the lower tangent altitudes, transitioning to a low bias in altitude at the upper tangent altitudes, (4) P_{HI} is consistently lower in pressure by 30–80 hPa than P_{AVG} (or 1–2 km), (5) the 1–σ variability is less for P_{HI} compared to P_{AVG}, which suggests that averaging multiple AIRS P_{CLD} retrievals along the MLS line of sight is less accurate than using the highest AIRS P_{CLD}, and (6) the differences in AIRS and MLS P_{CLD} are statistically significant at the highest MLS tangent altitudes, implying that AIRS and MLS "agree" at lower tangent altitudes, but "disagree" at higher tangent altitudes.

Finding (6) is consistent with the reduced sensitivity of AIRS to optically thin cirrus near the tropopause. The differences in AIRS and MLS P_{CLD} have contributions from sensitivity to particle size, sampling errors (e.g., using P_{AVG} or P_{HI} for AIRS P_{CLD}), and algorithmic differences. In Section (3), the comparisons of AIRS to ARM observations at Nauru Island showed that AIRS tends to place optically thin cirrus low in altitude. This behavior is consistent with IR-derived P_{CLD} located well below the cloud top when compared to lidar [*Holz et al*. 2006]. We anticipate that the future CloudSat [*Stephens et al.* 2002] and CALIPSO [*Winker et al.* 2002] spaceborne cloud radar and lidar missions will provide accurate cloud height and IWC information. This will aid in understanding the capability of passive IR- and MW-derived cloud height and IWC measurements.

5. Summary and Conclusions

We investigate the uncertainty in retrieved cloud top height (Z) from the Atmospheric Infrared Sounder (AIRS) and Advanced Microwave Sounding Unit (AMSU) observing suite located onboard the Earth Observing System (EOS) Aqua platform [*Aumann et al.* 2003]. We compare Z_A with coincident and independent measurements of Z derived from the micropulse lidar and millimeter-wave cloud radar (Z_{ARM}) at the Atmospheric Radiation Measurement (ARM) program sites of Nauru and Manus Islands [*Ackerman and Stokes* 2003], and the inferred Z derived from the vertically resolved Microwave Limb Sounder (MLS) ice water content (IWC) retrievals [*Wu et al.* 2006].

We use the Active Remotely-Sensed Cloud Locations (ARSCL) value-added product (VAP) [*Clothiaux et al.* 2000] to constrain the AIRS upper level cloud top height (Z_A) uncertainties. The ARSCL VAP combines the different strengths of the micropulse lidar,

millimeter-wave cloud radar, microwave radiometer, and laser ceilometer into a single cloud top and base product, allowing for a comprehensive database for clouds with varying heights and thicknesses, optical depths (τ), and hydrometeor characteristics. For a limited set of AIRS overpasses coincident with Manus and Nauru Islands (April–September 2003), only the cloud radar was in operation at Manus, and only the micropulse lidar was operational at Nauru. The primary findings of the AIRS–ARM comparisons are:

- AIRS is sensitive to a wide range of clouds, with statistically significant agreement for an effective cloud fraction (f_A) > 0.05.
- For all cases of $f_A > 0.15$ at Manus using the cloud radar only, the mean difference and 1–σ variability in $Z_A - Z_{ARM}$ ranges from –2.2 to 1.6 km ± 1.0 to 4.2 km. The range of values is a result of using different techniques to determine Z_{ARM}, as well as day/night differences.
- The differences between Z_A and Z_{ARM} are not a function of the length of the time window at Manus, but time window dependence is seen at Nauru for tenuous cirrus, implying sampling biases are a function of cloud opacity.
- The three height methods give consistently different results in $Z_A - Z_{ARM}$; these differences are larger (smaller) for smaller (larger) f_A.
- For the cases of $0.05 \leq f_A < 0.15$ at Nauru using the micropulse lidar only, the mean difference and 1–σ variability in the agreement is –1.1 to 2.1 km ± 3.0 to 4.5 km. This agreement is substantially improved over that using the cloud radar at Manus for this range of f_A.

The near simultaneous measurements of clouds by the AIRS and MLS instruments permit cross-examination of horizontally resolved cloud height and fraction fields with vertically

resolved ice water content (IWC). Comparisons of cloud top pressure (P_{CLD}) inferred from coincident AIRS and MLS IWC retrievals show that the mean difference depends upon the method for determining AIRS P_{CLD} (as with the ARM comparisons), the peak altitude and maximum value of MLS IWC, and f_A. The primary findings of the AIRS–MLS comparisons are:

- The bias and 1–σ variability of AIRS–MLS P_{CLD} decreases for increasing MLS IWC. The best agreement is seen when comparisons are limited to $f_A > 0.1$ and IWC > 1.0 mg m^{-3}.
- When using the highest AIRS P_{CLD} (P_{HI}) along the MLS tangent line of sight, the bias is lower in pressure by 30–80 hPa than when averaging the nearest three AIRS retrievals of P_{CLD} (P_{AVG}).
- The 1–σ variability is consistently lower for P_{HI} compared to P_{AVG}, suggesting that an average of multiple AIRS P_{CLD} retrievals along the MLS line of sight is often less accurate than using the highest AIRS P_{CLD}.
- Differences in AIRS and MLS P_{CLD} are statistically significant at the highest MLS tangent altitudes. This implies that AIRS and MLS statistically "agree" at lower tangent altitudes, but "disagree" at higher tangent altitudes. The disagreement occurs in regions dominated by tenuous cirrus near the tropopause.

Both AIRS and MLS have strengths and weaknesses in their abilities to sense cloud structure. The AIRS is capable of providing wide swaths of vertically integrated IR radiances, while the MLS is ideal for making measurements of clouds at discrete vertical intervals because of the limb viewing geometry. By cross-comparing the retrieved cloud quantities of instruments like AIRS and MLS (which view the same cloud fields from very different viewing perspectives)

to active cloud radar and lidar observations of clouds, a step towards characterizing the 3–D picture of upper tropospheric cloud structure is taken.

Appendix

This is a description of the test of the null hypothesis that observed agreement is due to randomness alone. Specifically, for each case in Table 1 (the different definitions of Z_{ARM} categorized by f_A), we begin with a set of bivariate data points (Z_{ARM}, Z_A), i=1,…, N, where N is the number of data point pairs. We measure the agreement between Z_{ARM} and Z_A by Δ, defined below as

$$\Delta = \frac{1}{N}\sum_{i=1}^{N}(Z_A - Z_{ARM})^2 \quad . \qquad \text{(A.1)}$$

The value Δ is the average of the squared differences between Z_A and Z_{ARM} on the L_2 norm, since the dynamic range of the variables is small [*Menke* 1989]. We calculate the true value of delta (Δ^*), as shown in Eqn (6), using N data point pairs. Then, we simulate the distribution of Δ under the assumption that Z_A and Z_{ARM} are *not* related. For each of B=10,000 trials in our simulation, we randomly re-ordered the Z_{ARM} values and computed Δ_b, for b=1, … , B. The histogram of Δ_b is an estimate of the so-called "null distribution" of Δ: the distribution under the assumption that the null hypothesis is true. The *p*-value of the hypothesis test is the proportion of Δ_b less than or equal to Δ^*. Typically, the null hypothesis is rejected if the *p*-value is less than 0.05 or, with a stricter standard, 0.01, corresponding to 5% and 1% confidence.

Resampling tests rely on assumptions that the data are a representative sample of the population from which they are drawn; in this case, the hypothetical population of Δ^* we would obtain were we able to examine all possible coincident Z_A and Z_{ARM}. Since the sample sizes (N) here are relatively small ($N < 20$ in most cases), the validity of these results is heavily dependent on the assumption of representativeness. There are (at least) three reasons to believe that the sample is reasonably representative: (1) each coincident case has been inspected individually and is seen to represent a wide variety of cloud scenes within each category of f_A, (2) the conclusions drawn from these results are consistent with the expected sensitivity of Z_A and Z_{ARM}, and (3) the histograms of the comparison compare favorably with those from other data sources. A longer time series of collocated measurements will likely give an increased level of representativeness.

Acknowledgments:

The authors thank John Blaisdell for detailed discussion on the AIRS cloud retrieval system. We thank all colleagues on the AIRS and MLS teams for support and helpful comments. BHK was funded by a National Research Council Resident Research Associate fellowship while in residence at NASA's Jet Propulsion Laboratory (JPL). Value-added product data were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate Change Research Division. This work was performed at JPL, California Institute of Technology, Pasadena, California, under contract with NASA.

References:

Ackerman, T.P., and G.M. Stokes (2003), The Atmospheric Radiation Measurement program, *Physics Today*, *56*, 38–44.

Aumann, H.H., M.T. Chahine, C. Gautier, M.D. Goldberg, E. Kalnay, L.M. McMillan, H. Revercomb, P.W. Rosenkranz, W.L. Smith, D.H. Staelin, L.L. Strow, and J. Susskind (2003), AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sensing*, *41*, 253–264.

Baum, B.A., and B.A. Wielicki (1994), Cirrus cloud retrievals using infrared sounding data: Multilevel cloud errors, *J. Appl. Meteor., 33*, 107–117.

Berendes, T.A., D.A. Berendes, R.M. Welch, E.G. Dutton, T. Uttal, and E.E. Clothiaux (2004), Cloud cover comparisons of the MODIS daytime cloud mask with surface instruments at the north slope of Alaska ARM site, *IEEE Trans. Geosci. Remote Sens., 42*, 2584–2593.

Cahalan, R.F., and J.H. Joseph (1989), Fractal statistics of cloud fields, *Mon. Wea. Rev*., *117*, 261–272.

Chahine, M.T. (1974), Remote sounding of cloudy atmospheres. I. The single cloud layer, *J. Atmos. Sci*., *31*, 233–243.

Chylek, P., and C. Borel (2004), Mixed phase cloud water/ice structure from high spatial

resolution satellite data, *Geophys. Res. Lett.*, *31*, L14104, doi:10.1029/2004GL020428.

Clothiaux, E.E. T.P. Ackerman, G.G. Mace, K.P. Moran, R.T. Marchand, M.A. Miller, and B.E. Martner (2000), Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites, *J. Appl. Meteor.*, *39*, 645–665.

Comstock, J.M., T.P. Ackerman, and G.G. Mace (2002), Ground-based lidar and radar remote sensing of tropical cirrus clouds at Nauru Island: Cloud statistics and radiative impacts, *J. Geophys. Res.*, *107*, 4714, doi:10.1029/2002JD002203.

Comstock, J.M., and C. Jakob (2004), Evaluation of tropical cirrus cloud properties derived from ECMWF model output and ground based measurements over Nauru Island, *Geophys. Res. Lett.*, *31*, L10106, doi:10.1029/2004GL019539.

Cooper, S.J., T.S. L'Ecuyer, and G.L. Stephens (2003), The impact of explicit cloud boundary information on ice cloud microphysical property retrievals from infrared radiances, *J. Geophys. Res., 108*, 4107, doi:10.1029/2002JD002611.

Cracknell, A.P. (1998), Synergy in remote sensing – what's in a pixel?, *Int. J. Remote Sensing, 19*, 2025–2047.

Cunningham, J.D., and J.M. Haas (2004) NPOESS instruments: The future of METSAT observations, In 11th Conference on Satellite Meteorology and Oceanography, Seattle, WA, American Meteorological Society.

Dessler, A.E., S.P. Palm, and J. D. Spinhirne (2006), Tropical cloud-top height distributions revealed by ICESat/GLAS, *J. Geophys. Res*., in press.

Elouragini, S., H. Chtioui, and P.H. Flamant (2005), Lidar remote sounding of cirrus clouds and comparison of simulated fluxes with surface and METEOSAT observations, *Atm. Res., 73*, 23–36.

Fetzer, E.J., B. Lambrigtsen, A. Eldering, H.H. Aumann, and M.T. Chahine (2006), Biases in total precipitable water vapor climatologies from AIRS and AMSR-E, *J.Geophys. Res*., in press.

Frey, R.A., B.A. Baum, W.P. Menzel, S.A. Ackerman, C.C. Moeller, and J.D. Spinhirne (1999), A comparison of cloud top heights computed from airborne lidar and MAS radiance data using CO_2 slicing, *J. Geophys. Res., 104*, 24,547–24,555.

Gettelman, A., W.D. Collins, E.J. Fetzer, A. Eldering, F.W. Irion, P.B. Duffy, and G. Bala (2006), Climatology of upper tropospheric relative humidity from the Atmospheric Infrared Sounder and implications for climate, *J. Climate*, in review.

Hawkinson, J.A., W. Feltz, and S.A. Ackerman (2005), A comparison of GOES sounder– and cloud lidar– and radar–retrieved cloud-top heights, *J. Appl. Meteor*., *44*, 1234–1242.

Heymsfield, A.J., and G.M. McFarquhar (2002), Mid-latitude and tropical cirrus: Microphysical properties, *Cirrus*, D.K. Lynch, K. Sassen, D. O'C. Starr, and G.L. Stephens, Eds., Oxford Press, 433–448.

Hollars, S., Q. Fu, J. Comstock, and T. Ackerman (2004), Comparison of cloud-top height retrievals from ground-based 35 Ghz MMCR and GMS-5 satellite observations at ARM TWP Manus site, *Atm. Res., 72*, 169–186.

Holz, R.E., S. Ackerman, P. Antonelli, F. Nagle, R.O. Knuteson, M. McGill, D.L. Hlavka, and W.D. Hart (2006), An improvement to the high spectral resolution CO_2 slicing cloud top altitude retrieval, *J. Atmos. Ocean. Tech*., in press.

Horváth, Á., and R. Davies (2004), Anisotropy of water cloud reflectance: A comparison of measurements and 1D theory, Geophys. Res. Lett., 31, L01102, doi:10.10292003GL018386.

Houghton, J. T., et al. (2001) Climate Change 2001: The Scientific Basis, Cambridge Univ. Press, New York, 881 pp.

Huang, H.–L., P. Yang, H. Wei, B.A. Baum, Y. Hu, P. Antonelli, and S.A. Ackerman (2004), Inference of ice cloud properties from high spectral resolution infrared observations, *IEEE Trans Geosci. Remote Sensing*, *42*, 842–853.

Jiang, J. H., D. L. Wu, H. Su, and J.W. Waters (2006), Ice Clouds in the Upper Troposphere as Observed by Microwave Limb Sounder on Aura Satellite, *J. Geophys. Res*., submitted.

Kahn, B.H., A. Eldering, F.W. Irion, F.P. Mills, B. Sen, and M.R. Gunson (2002), Cloud identification in Atmospheric trace Molecule Spectroscopy infrared occultation measurements, *Appl. Opt*., *41*, 2768–2780.

Kahn, B.H., K.N. Liou, S.-Y. Lee, E.F. Fishbein, S. DeSouza-Machado, A. Eldering, E.J. Fetzer, S.E. Hannon, and L.L. Strow (2005), Nighttime cirrus detection using Atmospheric Infrared Sounder window channels and total column water vapor, *J. Geophys. Res., 110*, doi:10.1029/2004JD005430.

Kahn, B.H., E. Fishbein, A. Eldering, E.J. Fetzer, M.J. Garay, and S.-Y. Lee (2006), A radiative consistency check between Atmospheric Infrared Sounder and Moderate Resolution Imaging Spectroradiometer cloud retrievals, to be submitted to *J. Geophys. Res.*

King, M.D., Y.J. Kaufman, W.P. Menzel, and D. Tanré (1992), Remote-sensing of cloud, aerosol, and water-vapor properties from the Moderate Resolution Imaging Spectroradiometer (MODIS), *IEEE Trans. Geosci. Remote Sensing*, *30*, 2–27.

Liou, K.N., S.C. Ou, Y. Takano, J. Roskovensky, G.G. Mace, K. Sassen, and M. Poellot (2002), remote sensing of three-dimensional inhomogeneous cirrus clouds using satellite and mm-wave cloud radar data, *Geophys. Res. Lett*., *29*, 1360, doi:10.1029/2002GL014846.

Livesey, N.J., W.V. Snyder, W.G. Read, P.A. Wagner (2006), Retrieval algorithms for the EOS Microwave Limb Sounder (MLS) instrument, *IEEE Trans. Geosci. Remote Sensing,* in press.

Mace, G.G., Y. Zhang, S. Platnick, M.D. King, P. Minnis, and P. Yang (2005), Evaluation of cirrus cloud properties derived from MODIS data using cloud properties derived from ground-based observations collected at the ARM SGP site, *J. Appl. Meteor., 40*, 221–240.

Mahesh, A., M.A. Gray, S.P. Palm, W.D. Hart, and J.D. Spinhirne (2004), Passive and active detection of clouds: Comparisons between MODIS and GLAS observations, *Geophys. Res. Lett*., *31*, L04108, doi:10.1029/2003GL018859.

Mather, J.H. (2005), Seasonal variability in clouds and radiation at the Manus ARM site, *J. Climate*, *18*, 2417–2428.

Menke, W. (1989), *Geophysical data analysis: Discrete inverse theory*, Academic Press, 289pp.

Menzel, W.P., W.L. Smith, and T.R. Stewart (1983), Improved cloud motion wind vector and altitude assignment using VAS, *J. Climate Appl. Meteor*., 22, 377–384.

Naud, C., J.-P. Muller, and E.E. Clothiaux (2002), Comparison of cloud top heights derived from MISR stereo and MODIS CO_2-slicing, *Geophys. Res. Lett., 29*, doi:10.1029/2002GL015460.

Naud, C., J.-P. Muller, M. Haeffelin, Y. Morille, and A. Delaval (2004), Assessment of MISR and MODIS cloud top heights through intercomparisons with a back-scattering lidar at SIRTA, *Geophys. Res. Lett., 31*, L04114, doi:10.1029/2003GL018976.

Peter, T. et al. (2004), Ultrathin Tropical Tropopause Clouds (UTTCs): I. Cloud morphology and occurrence, *Atmos. Chem. Phys*., *3*, 1083–1091.

Sherwood, S.C., J.-H. Chae, P. Minnis, and M. McGill (2004), Underestimation of deep convective cloud tops by thermal imagery, *Geophys. Res. Lett., 31*, L11102, doi:10.1029/2004GL019699.

Smith, W.L., and C.M.R. Platt (1978), Comparison of satellite-deduced cloud heights with indications from radiosonde and ground-based laser measurements, *J. Appl. Meteor., 17*, 1796–1802.

Smith, W.L., and R. Frey (1990), On cloud altitude determinations from high resolution interferometer sounder (HIS) observations, *J. Appl. Meteor., 29*, 658–662.

Spinhirne, J.D., S.P. Palm, W.D. Hart, D.L. Hlavka, and E.J. Welton (2005), Cloud and aerosol measurements from GLAS: Overview and initial results, *Geophys. Res. Lett*., *32*, L22S03, doi:10.1029/2005GL023507.

Stephens, G.L., D.G. Vane, R.J. Boain, G.G. Mace, K. Sassen, Z. Wang, A.J. Illingworth, E.J. O'Connor, W.B. Rossow, S.L. Durden, S.D. Miller, R.T. Austin, A. Benedetti, C. Mitrescu, and the CloudSat Science Team (2002), The CloudSat mission and the A-train, *Bull. Am. Met. Soc*., *83*, 1771–1790.

Stubenrauch, C.J., W.B. Rossow, N.A. Scott, and A. Chédin (1999), Clouds as seen by satellite sounders (3I) and imagers (ISCCP). Part III: Spatial heterogeneity and radiative effects, *J. Climate, 12*, 3419–3442.

Stubenrauch, C.J., F. Eddounia, and L. Sauvage (2005), Cloud heights from TOVS Path-B: Evaluation using LITE observations and distributions of highest cloud layers, *J. Geophys. Res*., *110*, doi:10.1029/2004JD005447.

Susskind, J., C.D. Barnett, and J.M. Blaisdell (2003), Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE Trans. Geosci. Remote Sens., 41*, 390–409.

Susskind, J., J. Blaisdell, L. Iredell, F. Keita, L. Kouvaris, G. Molnar, R. Atlas, and M. Chahine (2006), Accuracy of geophysical parameters derived from AIRS/AMSU as a function of fractional cloud cover, *J. Geophys. Res*., in press.

Vaubauce, C., B. Cadet, and R.T. Marchand (2003), Comparison of POLDER apparent and corrected oxygen pressure to ARM/MMCR cloud boundary pressures, *Geophys. Res. Lett., 30*, 1212, doi:10.1029/2002GL016449.

Waters, J.W., et al. (2006), The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Remote Sensing*, in press.

Wielicki, B.A., and J.A. Coakley, Jr. (1981), Cloud retrieval using infrared sounder data: error analysis, *J. Appl. Meteor., 20*, 157–169.

Winker, D.M., J. Pelon, and M.P. McCormick (2003), The CALIPSO mission: Spaceborne lidar for observation of aerosols and clouds, SPIE, 4893, Lidar Remote Sensing for Industry and Environment Monitoring III.

Wu, D.L., J.H. Jiang, and C.P. Davis (2006), EOS MLS cloud ice measurements and cloudy-sky radiative transfer model, *IEEE Trans. Geosci. Remote Sensing*, in press.

Wu, D.L., and J.H. Jiang (2004), EOS MLS algorithm theoretical basis for cloud measurements, JPL document D-19299, ATBD-MLS-06.

Wylie, D.P., and P.-H. Wang (1999), Comparison of SAGE-II and HIRS co-located cloud height measurements, *Geophys. Res. Lett., 26*, 3373–3375.

Zhang, H., and W.P. Menzel (2002), Improvement in thin cirrus retrievals using an emissivity-adjusted CO_2 slicing algorithm, *J. Geophys. Res., 107*, 4327, doi:10.1029/2001JD001037.

Figure Captions:

Figure 1. Comparisons of 78 AIRS and ARM cloud top height (Z_{CLD}) coincidences at Manus Island for nighttime granules. (a) AIRS and ARM Z_{CLD} using the AVG approach, (b) the HIST approach, and (c) the MAX approach; see text for details. Circles are proportional to upper level effective cloud fraction between 0.0 and 1.0. The gray scale represents the 960 cm^{-1} BT averaged over the AMSU footprint-scale: light gray to black is for a BT range of 300 – 190 K, respectively. Error bars for AIRS Z_{CLD} are the AIRS Standard L2 retrieval, and for the ARM Z_{CLD} they are the 1–σ level of Z_{CLD} variability over the 54 min time window.

Figure 2. Distributions of differences in Z_{CLD} between AIRS and ARM measurements for (a) AIRS–ARM AVG Z_{CLD} and (b) AIRS–ARM HIST Z_{CLD}, for the three time windows. Superimposed is AIRS–ARM MAX Z_{CLD} for the ± 24 min period, as well as an example of a "randomly-generated" agreement; see text for details on this experiment. All histograms are for the same 78 coincident observations presented Figure 1.

Figure 3. As Figure 1, except for coincident daytime AIRS granules at the Manus Island site.

Figure 4. As Figure 2 except for Manus Island daytime cases.

Figure 5. As Figure 1 for the coincident nighttime AIRS granules at the ARM Nauru Island site. The range of f is limited to 0.0–0.15 (increasing $f \rightarrow$ increasing diameter of circle); coincidences

with $f > 0.15$ are not included here. The three panels include the three different time-averaging periods; the AVG and MAX cases are not shown.

Figure 6. As Figure 2, except three time windows for the HIST case only. Top: all cases with $0.0 < f \leq 0.15$; middle: the subset of cases with $0.0 < f \leq 0.05$; the subset of cases with $0.05 < f \leq 0.15$.

Figure 7. Frequency histogram of the agreement between an active and passive-derived Z_{CLD} obtained from several independent data sources. (a) Comparison of ground-based radar with GMS-5 (*Hollars et al.* [2004], "All Clouds" in Figure 6) and AIRS Z_{CLD} (Figure 2, this paper). (b) Comparison of aircraft lidar and the MODIS Airborne Simulator Z_{CLD} (*Frey et al.* [1999], Figure 1), ground-based lidar+radar and GOES Z_{CLD} (*Hawkinson et al.* [2005], Figures 3 and 5 for the 3×3 and single FOV, respectively), and ground-based lidar and AIRS Z_{CLD} (Figure 6, this paper).

Figure 8. (a) Frequency distribution of AIRS P_{CLD} for collocated AIRS and MLS observations during 1 and 3–20 January 2005 within ±30° latitude (N=3726). All coincident cloud observations between AIRS and MLS are used; we do not consider cases where only one instrument detects clouds. Shown are the two AIRS P_{CLD} (AVG and HIGH) approaches (see text for details) for all values of upper cloud fraction (f_A), and $f_A > 0.1$. (b) Same as (a) except the frequency distribution is for the MLS P_{CLD}. The AIRS P_{CLD} (high) curve for $f_A > 0.1$ is repeated from (a) for comparison purposes. Also shown is the frequency distribution for cases excluding maximum MLS IWC < 1.0 mg m^{-3}, and the highest MLS level with IWC < 1.0 mg m^{-3}. The

MLS bins are centered on the tangent pressure altitudes from 68–215 hPa, while the AIRS bins are in 20 hPa intervals centered at 50, 70, …, 470, 490 hPa.

Figure 9. (a) Frequency distribution of AIRS (HIGH) – MLS P_{CLD} versus MLS pressure level using the same cases presented in Figure 8. (b) Same as (a) except for AIRS (AVG) P_{CLD}.

Table captions:

Table 1. The bias and 1–σ variability in km for $Z_A - Z_{ARM}$. Superimposed in italic (bold) are the *p*-values for a 1-sided hypothesis test for the 5% (1%) significance level where we reject the hypothesis that considers the agreement between $Z_A - Z_{ARM}$ to be due to random chance. There are three sets of observations: (1) Manus Island during daytime, (2) Manus Island during nighttime, and (3) Nauru Island during nighttime. The number of samples (*N*) is listed above each set of observations for each category of f_A.

Table 2. The AIRS–MLS difference in P_{CLD} (hPa): positive (negative) differences imply a higher (lower) AIRS mean pressure over its MLS counterpart. The same cases used in Figures 8 and 9 are presented here. The categories are partitioned by IWC. Column IWC value indicates the highest altitude with nonzero IWC in each MLS observation. The AIRS P_{CLD} is reported two ways: (1) the three nearest AIRS P_{CLD} retrievals relative to the MLS tangent point at 100 mb are averaged (P_{AVG}), and (2) the minimum AIRS P_{CLD} retrieval of the nearest three is used (P_{HI}). Bold and underlined numbers indicate that P_M and AIRS P_{CLD} are from statistically significant cloud distributions at the 5% significance level.

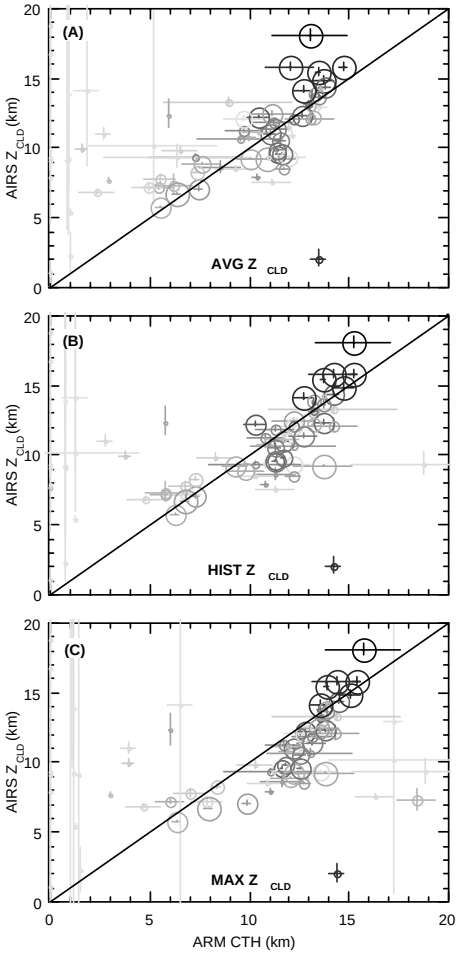

Figure 1. Comparisons of 78 AIRS and ARM cloud top height (Z_{CLD}) coincidences at Manus Island for nighttime granules. (a) AIRS and ARM Z_{CLD} using the AVG approach, (b) the HIST approach, and (c) the MAX approach; see text for details. Circles are proportional to upper level effective cloud fraction between 0.0 and 1.0. The gray scale represents the 960 cm^{-1} BT averaged over the AMSU footprint-scale: light gray to black is for a BT range of 300–190 K, respectively. Error bars for AIRS Z_{CLD} are the AIRS Standard L2 retrieval, and for the ARM Z_{CLD} they are the 1–σ level of Z_{CLD} variability over the 54 min time window.

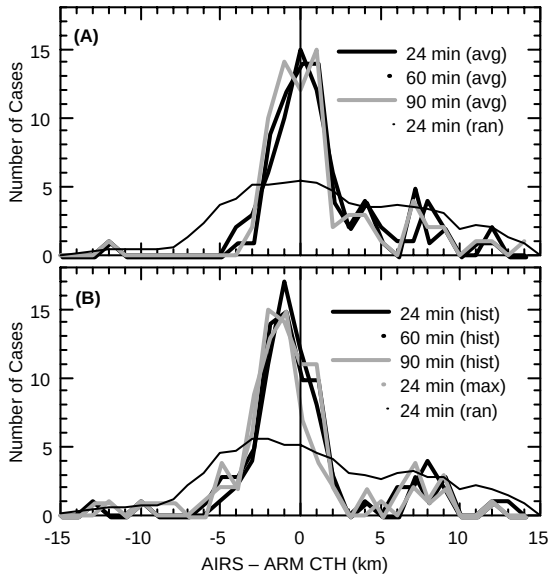

Figure 2. Distributions of differences in Z_{CLD} between AIRS and ARM measurements for (a) AIRS–ARM AVG Z_{CLD} and (b) AIRS–ARM HIST Z_{CLD}, for the three time windows. Superimposed is AIRS–ARM MAX Z_{CLD} for the ± 24 min period, as well as an example of a "randomly-generated" agreement; see text for details on this experiment. All histograms are for the same 78 coincident observations presented Figure 1.

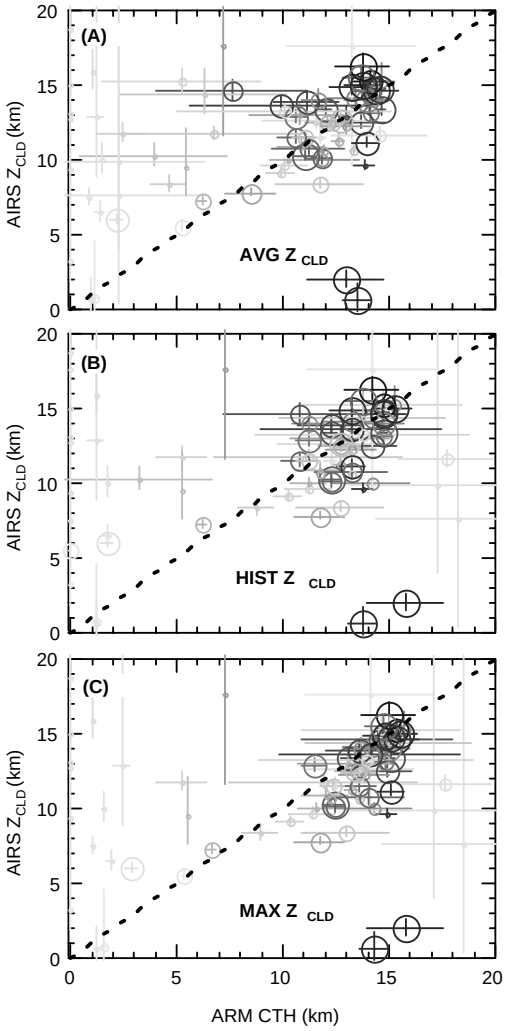

Figure 3. As Figure 1, except for coincident daytime AIRS granules at the Manus Island site.

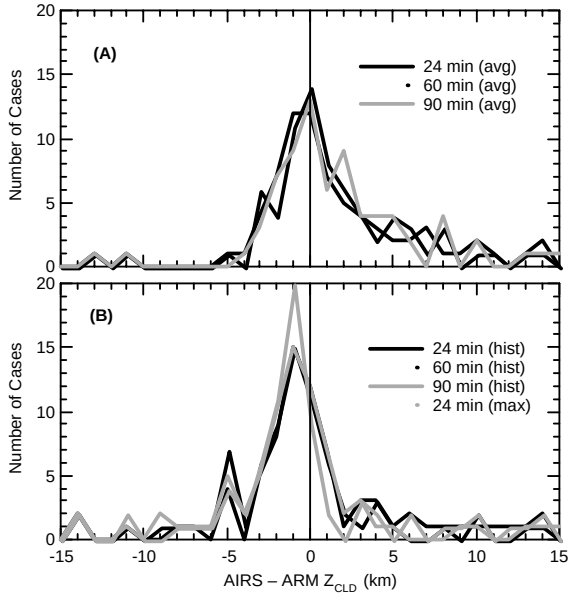

Figure 4. As Figure 2 except for Manus Island daytime cases.

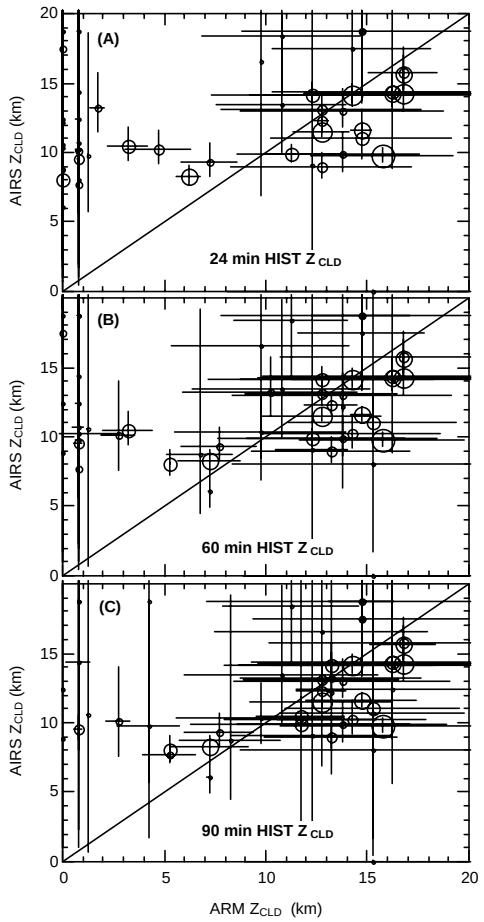

Figure 5. As Figure 1 for the coincident nighttime AIRS granules at the ARM Nauru Island site. The range of f is limited to 0.0–0.15 (increasing $f \rightarrow$ increasing diameter of circle); coincidences with $f > 0.15$ are not included here. The three panels include the three different time-averaging periods; the AVG and MAX cases are not shown.

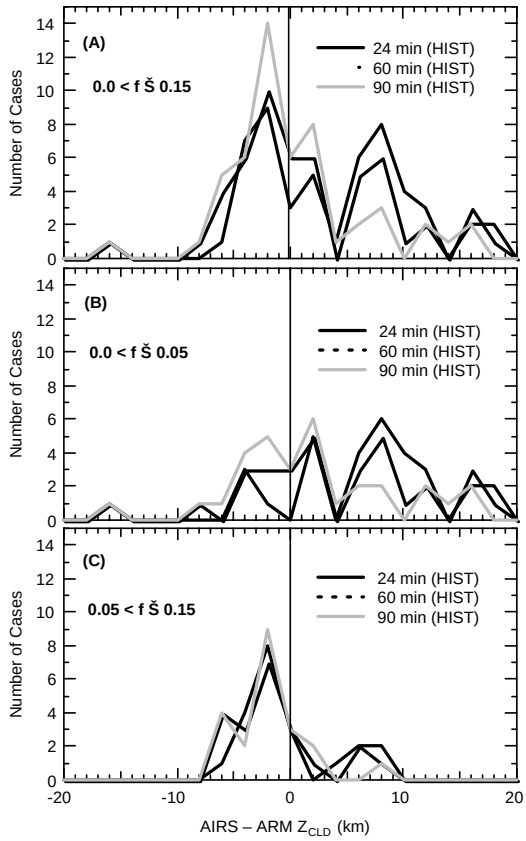

Figure 6. As Figure 2, except three time windows for the HIST case only. Top: all cases with $0.0 < f \leq 0.15$; middle: the subset of cases with $0.0 < f \leq 0.05$; the subset of cases with $0.05 < f \leq 0.15$.

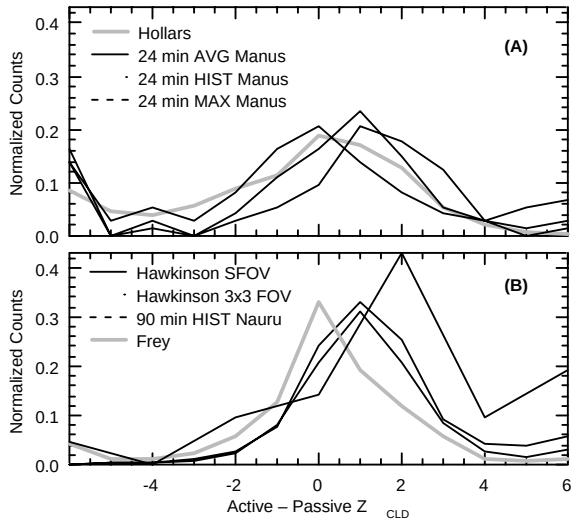

Figure 7. Frequency histogram of the agreement between an active and passive-derived Z_{CLD} obtained from several independent data sources. (a) Comparison of ground-based radar with GMS-5 (*Hollars et al.* [2004], "All Clouds" in Figure 6) and AIRS Z_{CLD} (Figure 2, this paper). (b) Comparison of aircraft lidar and the MODIS Airborne Simulator Z_{CLD} (*Frey et al.* [1999], Figure 1), ground-based lidar+radar and GOES Z_{CLD} (*Hawkinson et al.* [2005], Figures 3 and 5 for the 3×3 and single FOV, respectively), and ground-based lidar and AIRS Z_{CLD} (Figure 6, this paper).

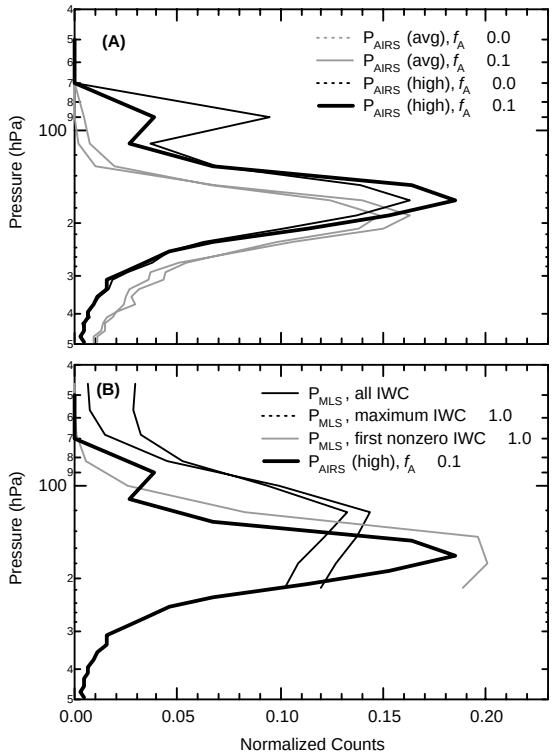

Figure 8. (a) Frequency distribution of AIRS P_{CLD} for collocated AIRS and MLS observations during 1 and 3–20 January 2005 within ±30° latitude (N=3726). All coincident cloud observations between AIRS and MLS are used; we do not consider cases where only one instrument detects clouds. Shown are the two AIRS P_{CLD} (AVG and HIGH) approaches (see text for details) for all values of upper cloud fraction (f_A), and $f_A > 0.1$. (b) Same as (a) except the frequency distribution is for the MLS P_{CLD}. The AIRS P_{CLD} (high) curve for $f_A > 0.1$ is repeated from (a) for comparison purposes. Also shown is the frequency distribution for cases excluding maximum MLS IWC < 1.0 mg m^{-3}, and the highest MLS level with IWC < 1.0 mg m^{-3}. The MLS bins are centered on the tangent pressure altitudes from 68–215 hPa, while the AIRS bins are in 20 hPa intervals centered at 50, 70, …, 470, 490 hPa.

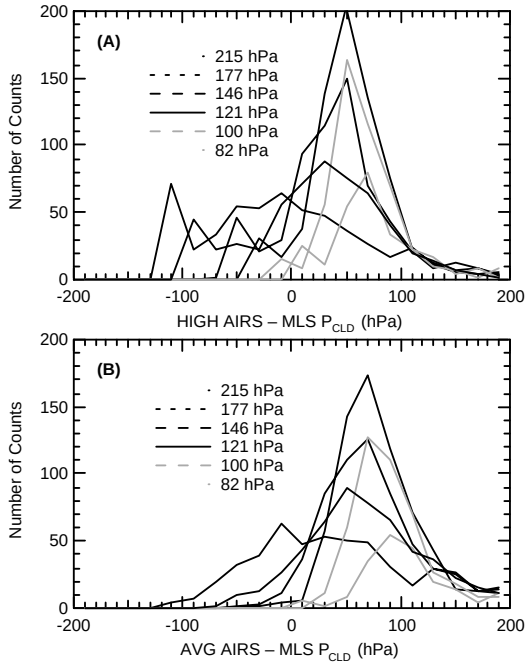

Figure 9. (a) Frequency distribution of AIRS (HIGH) – MLS P_{CLD} versus MLS pressure level using the same cases presented in Figure 8. (b) Same as (a) except for AIRS (AVG) P_{CLD}.

Location/Time	Time (min)	Height Method	0. ≤ *f* < .05	.05 ≤ *f* < .15	.15 ≤ *f* < .5	.5 ≤ *f* < .85	.85 ≤ *f* < 1.0
Manus/Night	–	–	*N*=13	*N*=9	*N*=21	*N*=16	*N*=16
	54	AVG	7.2 ± 7.0	2.1 ± 3.4	*0.4 ±3.7*	**–0.1 ± 1.5**	**0.7 ± 1.8**
	126	AVG	7.1 ± 6.5	*1.8 ±3.2*	*0.5 ±3.6*	**–0.3 ± 1.2**	**0.7 ± 2.0**
	186	AVG	7.0 ± 6.5	*1.9 ±3.0*	*0.4 ±3.6*	**–0.4 ± 1.3**	**0.6 ± 2.0**
	54	HIST	7.1 ± 7.3	1.1 ± 5.1	**–0.9 ± 3.4**	**–0.5 ± 1.3**	**–0.1 ± 1.7**
	126	HIST	4.9 ± 7.4	–0.5 ± 4.5	**–0.9 ± 3.4**	**–1.2 ± 1.0**	**–0.3 ± 2.0**
	186	HIST	4.7 ± 7.5	–0.4 ± 4.1	**–1.0 ± 3.3**	**–1.2 ± 1.0**	**–0.2 ± 2.0**
	54	MAX	5.3 ± 8.4	0.6 ± 4.9	*–2.2 ±4.0*	**–1.4 ± 1.3**	**–0.8 ± 1.9**
Manus/Day	_	_	*N*=21	*N*=12	*N*=16	*N*=12	*N*=16
	54	AVG	7.6 ± 5.6	6.3 ± 5.8	*1.2 ±4.2*	**0.2 ± 2.3**	**1.1 ± 1.6**
	126	AVG	7.8 ± 5.6	4.5 ± 4.9	*1.3 ±3.9*	**0.5 ± 2.3**	**1.3 ± 1.6**
	186	AVG	9.0 ± 5.0	4.4 ± 4.7	*1.5 ±3.8*	**0.7 ± 2.4**	**1.6 ± 1.7**
	54	HIST	6.4 ± 8.8	5.4 ± 6.1	**–0.4 ± 3.7**	**–0.1 ± 2.7**	**0.5 ± 1.6**
	126	HIST	3.7 ± 9.5	–1.0 ± 8.3	*–0.7 ±3.8*	**–1.1 ± 2.1**	**0.4 ± 1.6**
	186	HIST	1.5 ± 7.8	–1.5 ± 8.5	*–0.8 ±3.8*	**–1.1 ± 2.1**	**0.4 ± 1.5**
	54	MAX	4.8 ± 8.3	3.1 ± 8.1	**–0.7 ± 3.8**	**–1.5 ± 1.7**	**–0.2 ± 1.4**
Nauru/Night	_	_	*N*=32	*N*=20	_	_	_
	54	AVG	8.2 ± 6.1	**2.1 ± 3.9**	_	_	_
	126	AVG	7.1 ± 6.1	**1.9 ± 3.2**	_	_	_
	186	AVG	6.3 ± 5.4	**1.9 ± 3.0**	_	_	_
	54	HIST	7.4 ± 7.3	**0.3 ± 4.1**	_	_	_
	126	HIST	5.3 ± 7.8	**–0.7 ± 3.7**	_	_	_
	186	HIST	3.0 ± 7.3	**–1.1 ± 3.1**	_	_	_
	54	MAX	7.0 ± 7.5	**–0.5 ± 4.5**	_	_	_

Table 1. The bias and 1–σ variability in km for $Z_A - Z_{ARM}$. Superimposed in italic (bold) are the *p*-values for a 1-sided hypothesis test for the 5% (1%) significance level where we reject the hypothesis that considers the agreement between $Z_A - Z_{ARM}$ to be due to random chance. There are three sets of observations: (1) Manus Island during daytime, (2) Manus Island during nighttime, and (3) Nauru Island during nighttime. The number of samples (*N*) is listed above each set of observations for each category of f_A.

P_{MLS}	P_{AIRS}	All Cases	IWC > 1	IWC > 2	IWC > 5	IWC > 10	IWC > 20
215.4	AVG	61 ± 101 (*544*)	61 ± 101 (*544*)	61 ± 101 (*544*)	55 ± 96 (*499*)	39 ± 83 (*308*)	31 ± 68 (*68*)
	MAX	2 ± 93 (*544*)	2 ± 93 (*544*)	2 ± 93 (*544*)	–4 ± 87 (*499*)	–13 ± 70 (*308*)	–10 ± 59 (*68*)
177.8	AVG	87 ± 84 (*577*)	87 ± 84 (*577*)	84 ± 80 (*543*)	64 ± 62 (*216*)	56 ± 55 (*14*)	–
	MAX	33 ± 77 (*577*)	33 ± 77 (*577*)	30± 71 (*543*)	15 ± 55 (*216*)	4 ± 47 (*14*)	–
146.8	AVG	93 ± 85 (*649*)	83 ± 67 (*591*)	**77 ± 57 (294)**	–	–	–
	MAX	49 ± 72 (*649*)	42 ± 55 (*591*)	38 ± 43 (*294*)	–	–	–
121.2	AVG	**93 ± 69 (704)**	**85 ± 54 (310)**	**81 ± 40 (26)**	–	–	–
	MAX	59 ± 56 (*704*)	54 ± 42 (*310*)	48 ± 26 (*26*)	–	–	–
100.0	AVG	**115 ± 86 (498)**	**100 ± 66 (89)**	–	–	–	–
	MAX	74 ± 65 (*498*)	58 ± 24 (*89*)	–	–	–	–
82.5	AVG	**174 ± 134 (281)**	**189 ± 122 (22)**	–	–	–	–
	MAX	**101 ± 106 (281)**	**67 ± 33 (22)**	–	–	–	–

Table 2. The AIRS–MLS difference in P_{CLD} (hPa): positive (negative) differences imply a higher (lower) AIRS mean pressure over its MLS counterpart. The same cases used in Figures 8 and 9 are presented here. The categories are partitioned by IWC. Column IWC value indicates the highest altitude with nonzero IWC in each MLS observation. The AIRS P_{CLD} is reported two ways: (1) the three nearest AIRS P_{CLD} retrievals relative to the MLS tangent point at 100 mb are averaged (P_{AVG}), and (2) the minimum AIRS P_{CLD} retrieval of the nearest three is used (P_{HI}). Bold and underlined numbers indicate that P_M and AIRS P_{CLD} are from statistically significant cloud distributions at the 5% significance level.